Assignment 4

Introduction to Machine Learning
Prof. B. Ravindran

1. Which of the following are convex functions?
   (a) $f(x) = (\sum_{i=1}^{n} x_i^p)^{1/p}$, where $x \in R^n$ and $p \geq 0$
   (b) $f(x) = \log(\sum_{i=1}^{n} \exp x_i)$ where $x_i \in R^n$
   (c) $f(x) = \sum_{i=1}^{n} \sin(x_i)$, where $x \in R^n$
   (d) $f(x) = \sum_{i=1}^{n} x_i \log x_i$, where $x \in R^n$

2. We discussed two approaches to classification, one that learns the discriminant functions, and the other that is based on modeling hyperplanes. Which of these approaches is more suitable for multi-class problems and why?
   (a) Discriminant functions; because they allow for a probabilistic interpretation of the predictions.
   (b) Hyperplane methods; because they allow for a probabilistic interpretation of the predictions.
   (c) Discriminant functions; because an appropriate set of functions will allow us to efficiently disambiguate class predictions.
   (d) Hyperplane methods; because we can use basis expansion to transform the input to a space where class-boundaries are linear.

3. Consider the following optimization problem

$$\begin{array}{rcl} \min & x^2 + 1 & \\ \text{s.t.} & (x-2)(x-4) & \leq \; 0 \end{array}$$

   Select the correct options regarding this optimization problem.
   (a) Strong Duality holds
   (b) Strong duality doesn't hold.
   (c) The Lagrangian can be written as $L(x, \lambda) = (1+\lambda)x^2 - 6\lambda x + 1 + 8\lambda$
   (d) The dual objective will be $g(\lambda) = \frac{-9\lambda^2}{1+\lambda} + 1 + 8\lambda$

4. Which of the following is/are true about the Perceptron classifier?
   (a) It can learn a OR function
   (b) It can learn a OR function
   (c) The obtained separating hyperplane depends on the order in which the points are presented in the training process.
   (d) For a linearly separable problem, there exists some initialization of the weights which might lead to non-convergent cases.

5. Which of the following is/are true regarding an SVM?

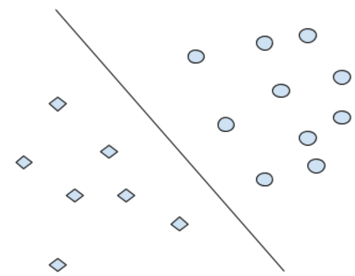

Figure 1: Q6

(a) For two dimensional data points, the separating hyperplane learnt by a linear SVM will be a straight line.

(b) In theory, a Gaussian kernel SVM can model any complex separating hyperplane.

(c) For every kernel function used in a SVM, one can obtain a equivalent closed form basis expansion.

(d) Overfitting in an SVM is a function of number of support vectors.

6. Consider a two class problem, whose training points are distributed in the figure below. One possible separating hyperplane is shown in the figure.

   (a) A classifier can be learnt using the perceptron training algorithm.

   (b) A linear SVM will not work well.

   (c) A linear SVM is sufficient for this data.

   (d) A non zero C value is essential for this data.

7. For a two-class classification problem, we use an SVM classifier and obtain the following separating hyperplane. We have marked 4 instances of the training data. Identify the point which will have the most impact on the shape of the boundary on it's removal.

   (a) 1

   (b) 2

   (c) 3

   (d) 4

8. For the dataset 1, train linear and radial basis function kernel SVMs. What are the number of Support Vectors in each of the case?

   (a) 100, 100

   (b) 10, 105

   (c) 3, 104

   (d) 500, 50

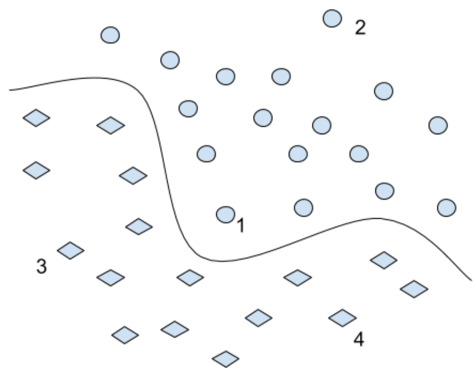

Figure 2: Q7

9. For dataset 2, train 5 degree polynomial (5 degree, coef0 = 0), 10 degree polynomial (10 degree, coef0 = 0) and radial basis kernel functions. What are the number of support vectors for each?
   (a) 10, 300, 56
   (b) 324, 20, 27
   (c) 43, 98, 76
   (d) 12, 27, 20
10. Based on the previous experiments, which would you think is the ideal classifier for Dataset 1.
   (a) Linear SVM
   (b) Polynomial SVM
   (c) Radial basis SVM